

LMS IS  $H^\infty$  OPTIMAL \*BABAK HASSIBI, ALI H. SAYED<sup>†</sup> AND THOMAS KAILATH*Information Systems Laboratory, Stanford University, CA 94305, USA.*

## ABSTRACT

We show that the celebrated LMS (Least-Mean Squares) adaptive algorithm is an  $H^\infty$  optimal filter. In other words, the LMS algorithm, which has long been regarded as an approximate least-mean squares solution, is in fact a minimizer of the  $H^\infty$  error norm and not the  $H^2$  norm. In particular, the LMS minimizes the energy gain from the disturbances to the *predicted* errors, while the normalized LMS minimizes the energy gain from the disturbances to the *filtered* errors. Moreover, since these algorithms are *central*  $H^\infty$  filters, they are also risk-sensitive optimal and minimize a certain exponential cost function. We discuss various implications of these results, and show how they provide theoretical justification for the widely observed excellent robustness properties of the LMS filter.

## I. INTRODUCTION

The LMS algorithm was originally conceived as an approximate recursive procedure that solves the following adaptive problem [1]: given a sequence of  $1 \times M$  input row vectors  $\{h_i\}$ , and a corresponding sequence of desired responses  $\{d_i\}$ , find an estimate of an  $M \times 1$  column vector of weights  $w$  such that the squared error sum  $\sum_{i=0}^N |d_i - h_i w|^2$  is minimized. The LMS solution recursively updates estimates of the weight vector along the direction of the instantaneous gradient of the squared error. Exact recursive least-squares (RLS) algorithms have also been developed (see, e.g., [2]). Although these have better convergence properties, they are computationally more complex, and exhibit poorer robust behaviour than the simple LMS.

<sup>†</sup>Dr. Sayed is currently with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, USA. This work was performed while he was a Research Associate in the Information Systems Laboratory, Stanford University, and on Leave from Escola Politécnica da Universidade de São Paulo, Brazil.

\* This work was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command under Contract AFOSR91-0060 and by the Army Research Office under contract DAAL03-89-K-0109.

More recently, and motivated by applications in control theory, there has been an increasing interest in  $H^\infty$ -filtering (see, e.g., [3]-[7] and the references therein) with the belief that the resulting minimax algorithms will be more robust and less sensitive to parameter variations, to model uncertainties, and to the lack of statistical information on the exogenous signals. In this paper we show that the LMS filter is a minimax algorithm, which provides a theoretical justification for its superior robust properties. More specifically, we shall use some of the results developed in the companion papers [8, 9] in order to show that the LMS algorithm is the central *a priori*  $H^\infty$ -optimal filter, while the so-called normalized LMS algorithm is the central *aposteriori*  $H^\infty$ -optimal filter. This provides a minimization criterion that has long been missing for the LMS algorithm. Moreover, since LMS and normalized LMS are shown here to be *central* filters they are also risk-sensitive optimal and minimize a certain exponential cost function [11].

II. THE  $H^\infty$  PROBLEM

We first give a brief review of some of the results described in the companion papers [8, 9] on  $H^\infty$ -filtering. The reader is also referred to [3]-[7], and the references therein, for earlier results and alternative approaches. We begin with the definition of the  $H^\infty$  norm of a transfer operator. As will presently become apparent, the motivation for introducing the  $H^\infty$  norm is to capture the worst case behaviour of a system.

Let  $h_2$  denote the vector space of square summable causal sequences  $\{f_k, 0 \leq k < \infty\}$ , with inner product  $\langle \{f_k\}, \{g_k\} \rangle = \sum_{k=0}^{\infty} f_k^* g_k$ , where  $*$  denotes complex conjugation. Let  $T$  be a transfer operator that maps a causal input sequence  $\{u_k\}$  to a causal output sequence  $\{y_k\}$ . The  $H^\infty$  norm of  $T$  is equal to

$$\|T\|_\infty = \sup_{u \in h_2, u \neq 0} \frac{\|y\|_2}{\|u\|_2},$$

where the notation  $\|u\|_2$  denotes the  $h_2$ -norm of the causal sequence  $\{u_k\}$ , viz.,  $\|u\|_2^2 = \sum_{k=0}^{\infty} u_k^* u_k$ .

The  $H^\infty$  norm can thus be regarded as the maximum *energy gain* from the input  $u$  to the output  $y$ .

## II.1 Problem Statement

We consider a state-space model of the form

$$\begin{aligned} x_{i+1} &= F_i x_i + G_i u_i, & x_0 \\ y_i &= H_i x_i + v_i, \end{aligned} \quad (1)$$

where  $x_0$ ,  $\{u_i\}$ , and  $\{v_i\}$  are *unknown* quantities and  $y_i$  is the measured output. Let  $z_i$  be linearly related to the state  $x_i$  via a given matrix  $L_i$ , viz.,  $z_i = L_i x_i$ .

We shall be interested in the following two problems. Let  $\hat{z}_{i|i} = \mathcal{F}_f(y_0, y_1, \dots, y_i)$  denote the estimate of  $z_i$  given observations  $\{y_j\}$  from time 0 up to and including time  $i$ , according to a certain error criterion to be made precise ahead, and let  $\hat{z}_i = \mathcal{F}_p(y_0, y_1, \dots, y_{i-1})$  denote the estimate of  $z_i$  given observations  $\{y_j\}$  from time 0 to time  $i-1$ . This defines two estimation errors: the *filtered* error  $e_{f,i} = \hat{z}_{i|i} - L_i x_i$ , and the *predicted* error  $e_{p,i} = \hat{z}_i - L_i x_i$ .

Let  $T_f$  ( $T_p$ ) denote the transfer operator that maps the unknown disturbances  $\{\Pi_0^{-1/2}(x_0 - \hat{x}_0), u_i, v_i\}$  to the filtered (predicted) error  $e_{f,i}$  ( $e_{p,i}$ ), where  $\hat{x}_0$  denotes an initial guess of  $x_0$  and  $\Pi_0$  denotes a positive definite matrix that reflects a priori knowledge as to how close  $x_0$  is to the initial guess  $\hat{x}_0$ . The  $H^\infty$  estimation problem(s) can now be stated as follows.

**Optimal  $H^\infty$  Problem.** Find  $H^\infty$  optimal estimation strategies  $\hat{z}_{i|i} = \mathcal{F}_f(y_0, y_1, \dots, y_i)$  and  $\hat{z}_i = \mathcal{F}_p(y_0, y_1, \dots, y_{i-1})$  that respectively minimize  $\|T_f\|_\infty$  and  $\|T_p\|_\infty$ , and obtain the resulting  $\gamma_{f,o}^2 = \inf_{\mathcal{F}_f} \|T_f\|_\infty^2$ , and  $\gamma_{p,o}^2 = \inf_{\mathcal{F}_p} \|T_p\|_\infty^2$ , where  $\gamma_{f,o}^2 =$

$$\inf_{\mathcal{F}_f} \left( \sup_{\substack{x_0 \\ u \in h_2 \\ v \in h_2}} \frac{\|e_{f,i}\|_2^2}{(x_0 - \hat{x}_0)^* \Pi_0^{-1} (x_0 - \hat{x}_0) + \|u\|_2^2 + \|v\|_2^2} \right)$$

and  $\gamma_{p,o}^2 =$

$$\inf_{\mathcal{F}_p} \left( \sup_{\substack{x_0 \\ u \in h_2 \\ v \in h_2}} \frac{\|e_{p,i}\|_2^2}{(x_0 - \hat{x}_0)^* \Pi_0^{-1} (x_0 - \hat{x}_0) + \|u\|_2^2 + \|v\|_2^2} \right)$$

The distinction between the *strictly* causal  $\mathcal{F}_p$  and the causal  $\mathcal{F}_f$  is significant since the solution to the  $H^\infty$  problem, as we shall see, depends on the

structure of the information available to the estimator. We can also infer from the above problem that the robust behaviour of  $H^\infty$  optimal estimators is because they guarantee the smallest estimation error energy over all possible disturbances of fixed energy.

A closed form solution of the optimal  $H^\infty$  problem is available only for some special cases (one of which is the adaptive filtering problem to be studied here), and a simpler problem results if one relaxes the minimization condition and settles for a suboptimal solution.

**Sub-optimal  $H^\infty$  Problem.** Given scalars  $\gamma_f > 0$  and  $\gamma_p > 0$ , find estimation strategies  $\hat{z}_{i|i} = \mathcal{F}_f(y_0, y_1, \dots, y_i)$  and  $\hat{z}_i = \mathcal{F}_p(y_0, y_1, \dots, y_{i-1})$  that respectively achieve  $\|T_f\|_\infty \leq \gamma_f$  and  $\|T_p\|_\infty \leq \gamma_p$ . This clearly requires checking whether  $\gamma_f \geq \gamma_{f,o}$  and  $\gamma_p \geq \gamma_{p,o}$ .

To guarantee  $\|T_f\|_\infty \leq \gamma_f$  we shall proceed as follows: let  $T_{f,i}$  be the transfer operator that maps  $\{\Pi_0^{-1/2}(x_0 - \hat{x}_0), \{u_j\}_{j=0}^i, \{v_j\}_{j=0}^i\}$  to the filtered errors  $\{e_{f,j}\}_{j=0}^i$ . We shall find a  $\gamma_f$  that ensures  $\|T_{f,i}\|_\infty < \gamma_f$  for all  $i$ . Likewise we shall find a  $\gamma_p$  that ensures  $\|T_{p,i}\|_\infty < \gamma_p$  for all  $i$ .

## III. THE $H^\infty$ -FILTERS

We now briefly review some of the results on  $H^\infty$  filters using the notation of [8, 9].

**Theorem 1 (Apriori Filter)** For a given  $\gamma > 0$ , if the  $F_j$  are nonsingular (for  $j \leq i$ ), then an estimator with  $\|T_{f,i}\|_\infty < \gamma$  exists iff

$$P_j^{-1} + H_j^* H_j - \gamma^{-2} L_j^* L_j > 0, \quad j = 0, \dots, i \quad (2)$$

where  $P_0 = \Pi_0$  and  $P_j$  satisfies the Riccati recursion

$$P_{j+1} = F_j(P_j^{-1} + H_j^* H_j - \gamma^{-2} L_j^* L_j)^{-1} F_j^* + G_j G_j^* \quad (3)$$

If this is the case, then one possible  $H^\infty$  filter with level  $\gamma$  is given by  $\hat{z}_{j|i} = L_j \hat{x}_{j|i}$ , where  $\hat{x}_{j|i}$  is recursively computed as follows:  $\hat{x}_{-1|-1} = \hat{x}_0$ ,

$$\hat{x}_{j+1|j+1} = F_j \hat{x}_{j|j} + K_{f,j}(y_{j+1} - H_{j+1} F_j \hat{x}_{j|j}),$$

where  $K_{f,j} = P_{j+1} H_{j+1}^* (I + H_{j+1} P_{j+1} H_{j+1}^*)^{-1}$ .

**Theorem 2 (Apriori Filter)** For a given  $\gamma > 0$ , if the  $F_j$  are nonsingular (for  $j \leq i$ ), an estimator with  $\|T_{p,i}\|_\infty < \gamma$  exists iff

$$\tilde{P}_j^{-1} = P_j^{-1} - \gamma^{-2} L_j^* L_j > 0, \quad j = 0, \dots, i \quad (4)$$

where  $P_j$  is the same as in Theorem 1. If this is the case, then one possible  $H^\infty$  filter with level  $\gamma$  is given by  $\hat{z}_j = L_j \hat{x}_j$ , where

$$\hat{x}_{j+1} = F_j \hat{x}_j + K_{p,j}(y_j - H_j \hat{x}_j), \quad \hat{x}_0$$

and  $K_{p,j} = \tilde{P}_j H_j^* (I + H_j \tilde{P}_j H_j^*)^{-1}$ .

Note that the above two estimators bear a striking resemblance to the celebrated Kalman filter, and that the major difference is that the  $P_j$  and  $\tilde{P}_j$  satisfy Riccati recursions that differ from that associated with the Kalman filter. Also, as  $\gamma \rightarrow \infty$  the Riccati recursion (3) collapses to the Kalman filter Riccati recursion, suggesting that the  $H^\infty$  norm of the Kalman filter may be quite large. It is also interesting to note that, contrary to the Kalman filter, the structure of the  $H^\infty$  estimators depends, via the Riccati recursion (3), on the linear combination of the states that we intend to estimate (i.e. the  $L_i$ ). Intuitively, this means that the  $H^\infty$  filters are specifically tuned towards the linear combination  $L_i x_i$ .

The filters of Theorems 1 and 2 are among many possible filters with level  $\gamma$ . Their full parametrization is given as follows (see also [9]).

**All Aposteriori Filters.** All  $H^\infty$  aposteriori estimators that achieve a level  $\gamma$  (assuming they exist) are given by

$$\begin{aligned} \hat{z}_{i|i} &= L_i \hat{x}_{i|i} + [I - L_i(P_i^{-1} + H_i^* H_i)^{-1} L_i^*]^{\frac{1}{2}} \\ &\quad S_i \left( \{(I + H_j P_j H_j^*)^{\frac{1}{2}} (y_j - H_j \hat{x}_{j|j})\}_{j=0}^i \right) \end{aligned}$$

where  $\hat{x}_{i|i}$  is given by Theorem 1, and  $S(a_i, \dots, a_0)$  is any (possibly nonlinear) contractive causal mapping of the form

$$S = \begin{bmatrix} S_0(a_0) \\ S_1(a_1, a_0) \\ \vdots \\ S_i(a_i, \dots, a_0) \end{bmatrix},$$

and satisfies  $\sum_{j=0}^i |S_j(a_j, \dots, a_0)|^2 \leq \sum_{j=0}^i |a_j|^2$ .

**All Apriori Filters.** All  $H^\infty$  apriori estimators that achieve a level  $\gamma$  (assuming they exist) are given by

$$\begin{aligned} \hat{z}_i &= L_i \hat{x}_i + [I - L_i P_i L_i^*]^{\frac{1}{2}} \\ &\quad S_i \left( \{(I + H_j \tilde{P}_j H_j^*)^{-\frac{1}{2}} (y_j - H_j \hat{x}_j)\}_{j=0}^{i-1} \right) \end{aligned}$$

where  $\hat{x}_i$  and  $\tilde{P}_i$  are given by Theorem 2, and  $S$  is any (possibly nonlinear) contractive causal mapping as above.

The filters of Theorems 1 and 2 are known as the central filters and correspond to  $S = 0$ .

#### IV. THE $H^\infty$ -ADAPTIVE PROBLEM

Suppose we observe an output sequence  $\{d_i\}$  that obeys the following model:

$$d_i = h_i w + v_i \quad (5)$$

where  $h_i = [h_1(i) \ h_2(i) \ \dots \ h_M(i)]$  is a known  $1 \times M$  input vector,  $w$  is an unknown  $M \times 1$  weight vector, and  $v_i$  is an unknown disturbance, which may also include modelling errors. We shall not make any assumptions on the noise sequence  $\{v_i\}$ , such as stationarity, whiteness, normal distributed, etc. Equation (5) can be reformulated into a state-space model as follows:

$$\begin{aligned} x_{i+1} &= x_i, & x_0 &= w \\ d_i &= h_i x_i + v_i. \end{aligned} \quad (6)$$

This is a relevant step since it reduces the adaptive filtering problem to an equivalent state-space estimation problem. This point of view has been recently used in [10] where a unified square-root-based derivation of exponentially-weighted RLS adaptive algorithms is obtained by reformulating the original adaptive problem as a state-space linear least-squares estimation problem and then applying various algorithms from Kalman filter theory. Here we shall instead apply the  $H^\infty$  filters to the state-space model (6) and show that they collapse to the LMS and normalized LMS algorithms.

Consider the uncorrupted output  $z_i = h_i x_i$  of (6). As before, define the estimates  $\hat{z}_{i|i}$  and  $\hat{z}_i$ , the estimation errors  $e_{f,i}$  and  $e_{p,i}$ , and the transfer operators  $T_f$  and  $T_p$ .

**$H^\infty$  Adaptive Problem.** Find  $H^\infty$ -optimal estimation strategies  $\hat{z}_{i|i} = \mathcal{F}_f(d_0, d_1, \dots, d_i)$  and  $\hat{z}_i = \mathcal{F}_p(d_0, d_1, \dots, d_{i-1})$  that respectively minimize  $\|T_f\|_\infty$  and  $\|T_p\|_\infty$ , and obtain the resulting

$$\gamma_{f,o}^2 = \inf_{\mathcal{F}_f} \|T_f\|_\infty^2 = \inf_{\mathcal{F}_f} \sup_{w, v \in h_2} \frac{\|e_f\|_2^2}{\mu^{-1} |w - \hat{w}_{|-1}|^2 + \|v\|_2^2} \quad (7)$$

and

$$\gamma_{p,o}^2 = \inf_{\mathcal{F}_p} \|T_p\|_\infty^2 = \inf_{\mathcal{F}_p} \sup_{w, v \in h_2} \frac{\|e_p\|_2^2}{\mu^{-1} |w - \hat{w}_{|-1}|^2 + \|v\|_2^2},$$

where  $\mu$  is a constant that reflects apriori knowledge as to how close  $w$  is to the initial guess  $\hat{w}_{|-1}$ .

#### V. MAIN RESULT

We now show that if we specialize the recursions of the apriori and aposteriori  $H^\infty$  filters to the state-space model (6), then the LMS and the normalized

LMS algorithms readily follow. We first add the assumption that the input vectors  $h_i$  are exciting, that is  $\lim_{N \rightarrow \infty} \sum_{i=0}^N h_i h_i^* = \infty$ .

**Theorem 3 (Normalized LMS)**

Consider the state-space model (6), and suppose we want to minimize the  $H^\infty$  norm of the transfer operator  $T_{f,i}$  from the unknowns  $w$  and  $\{v_j\}_{j=0}^i$  to the filtered error  $\{e_{f,j} = \hat{z}_{j|j} - h_j w\}_{j=0}^i$ . If the input data  $\{h_j\}$  is exciting, then the minimum  $H^\infty$  norm is  $\gamma_{opt} = 1$ . In this case, the central optimal  $H^\infty$  a posteriori filter is  $\hat{z}_{j|j} = h_j \hat{w}_{j|j}$ , where  $\hat{w}_{j|j}$  is given by the normalized LMS algorithm with parameter  $\mu$ :

$$\hat{w}_{j+1} = \hat{w}_{j|j} + \frac{\mu h_{j+1}^*}{1 + \mu h_{j+1} h_{j+1}^*} (d_{j+1} - h_{j+1} \hat{w}_{j|j}).$$

Intuitively it is not hard to convince oneself that  $\gamma_{opt}$  cannot be less than one. To this end, suppose that the estimator has chosen some initial guess  $\hat{w}_{|-1}$ . Then one may conceive of a disturbance that yields an observation that coincides with the output expected from  $\hat{w}_{|-1}$ , i.e.,  $h_i \hat{w}_{|-1} = h_i w + v_i = d_i$ . In this case one expects that the estimator will not change its estimate of  $w$ , so that  $\hat{w}_{|i} = \hat{w}_{|-1}$  for all  $i$ . Thus, the filtered error is  $e_{f,i} = h_i \hat{w}_{|i} - h_i w = h_i \hat{w}_{|-1} - h_i w = v_i$ , and the ratio in (7) can be made arbitrarily close to one.

**Theorem 4 (LMS)** If we instead want to minimize the  $H^\infty$  norm of the transfer operator  $T_{p,i}$  from the unknowns  $w$  and  $\{v_j\}_{j=0}^i$  to the predicted error  $\{e_{p,j} = \hat{z}_j - h_j w\}_{j=0}^i$ , and assuming the input data  $\{h_j\}$  is exciting and

$$0 < \mu < \inf_i (1/h_i h_i^*),$$

then the minimum  $H^\infty$  norm is  $\gamma_{opt} = 1$ . Moreover, the central optimal a priori  $H^\infty$  filter is  $\hat{z}_j = h_i \hat{w}_{j|-1}$ , where  $\hat{w}_{j|-1}$  is given by the LMS algorithm with learning rate  $\mu$ , viz.,

$$\hat{w}_{|j} = \hat{w}_{j|-1} + \mu h_j^* (d_j - h_j \hat{w}_{j|-1}).$$

The above result gives a bound on the learning rate  $\mu$  in order to guarantee the  $H^\infty$  optimality of LMS, which is in agreement with the well known fact that LMS behaves poorly if the learning rate is chosen too large.

We remark that if the input data is not exciting, then the LMS and normalized LMS algorithms will still correspond to  $\gamma = 1$ , but will now be suboptimal.

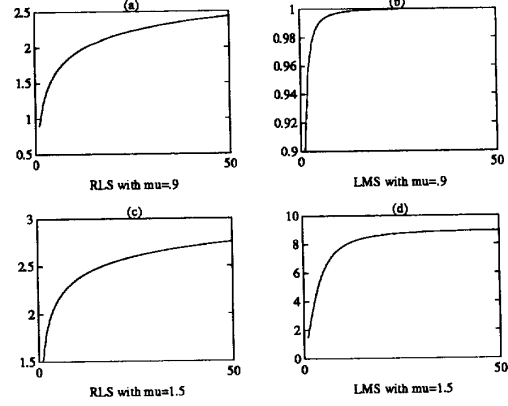


Figure 1: Maximum singular value of the transfer operators  $T_{lms,N}(\mu)$  and  $T_{rls,N}(\mu)$  as a function of  $N$ , for  $\mu = .9$  and  $\mu = 1.5$ .

### V.3 A Simulation

We have thus seen that LMS and normalized LMS ensure that the energy of the estimation errors never exceeds that of the disturbances. This is not true for other estimators, such as the recursive least-squares (RLS) algorithm [2], as we illustrate by an example.

For this purpose, we consider a special case of model (6) where  $h_i$  is taken as a scalar that randomly takes on the values  $+1$  and  $-1$ . In this special problem,  $\mu$  must be less than one to guarantee the  $H^\infty$  optimality of LMS. We thus choose the two values  $\mu = .9$  and  $\mu = 1.5$  (one greater and one less than  $\mu = 1$ ). The results are illustrated in Figure 1 where the maximum singular values of  $T_{lms,N}(\mu)$  and  $T_{rls,N}(\mu)$  (the transfer operators from the disturbances to prediction errors for LMS and RLS, respectively) are plotted against the number of observations  $N$ . As expected, for  $\mu = .9$  the maximum singular value of  $T_{lms,N}(\mu)$  remains constant at one, whereas the maximum singular value of  $T_{rls,N}(\mu)$  is greater than one and increases with  $N$ . For  $\mu = 1.5$  both RLS and LMS display maximum singular values greater than one.

The worst case RLS and LMS disturbance signals are found by computing the maximum singular vectors of  $T_{rls,50}(.9)$  and  $T_{lms,50}(.9)$ , and are shown in Figure 2. As can be seen from Figures 2b and 2d the LMS predicted error goes to zero while the RLS predicted error does not (in Figure 2b it is actually biased). The worst case disturbances (especially for RLS) are interesting; they compete with the true

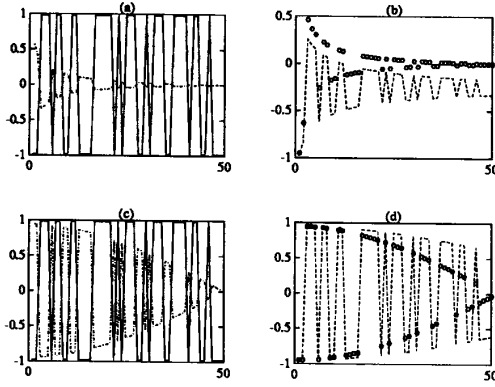


Figure 2: Worst case disturbances, and the corresponding predicted errors, for RLS and LMS in the  $\mu = .9$  case. (a) Solid line represents the uncorrupted output  $h_i x_i$  and the dashed line represents the worst case RLS disturbance. (b) The dashed line and the dotted line represent the RLS and LMS predicted errors, respectively, for the worst case RLS disturbance. (c) Solid line represents the uncorrupted output  $h_i x_i$  and the dashed line represents the worst case LMS disturbance. (d) The dashed line and the dotted line represent the RLS and LMS predicted errors, respectively, for the worst case LMS disturbance.

output early on, and then go to zero.

#### V.4 Discussion

In the beginning of this section we motivated the  $\gamma_{opt} = 1$  result for normalized LMS by considering a disturbance strategy that made the observed output  $d_i$  coincide with the expected output  $h_i \hat{w}_{i-1}$ . It is now illuminating to consider the *dual* strategy for the estimator. Such a strategy would be to construct an estimate that coincides with the observed output, *viz.*,

$$\hat{z}_{i|i} = d_i \quad (8)$$

The corresponding filtered error is  $e_{f,i} = \hat{z}_{i|i} - h_i x_i = d_i - h_i x_i = v_i$  so that the ratio in (7) can be made arbitrarily close to one, and the estimator (8) will achieve the same  $\gamma_{opt} = 1$  that the normalized LMS algorithm does. The fact that the simplistic estimator (8) (which is obviously of no practical use) is an optimal  $H^\infty$  aposteriori filter seems to question the very merit of being  $H^\infty$  optimal. The point is that  $H^\infty$  estimators that achieve a certain level  $\gamma$  are nonunique, and while the property of being  $H^\infty$  optimal may be desirable in several instances, different estimators in the set of all  $H^\infty$  optimal estima-

tors may have drastically different behaviour with respect to other desirable performance measures.

The LMS and the normalized LMS algorithms correspond to the so-called central filters. These are also risk-sensitive optimal filters, *i.e.*, they meet a certain exponential cost criterion, and can also be shown to be maximum entropy.

### VI. ALL $H^\infty$ -ADAPTIVE FILTERS

Using the parametrization theorems of Section III, we can parametrize all optimal  $H^\infty$  apriori and aposteriori adaptive filters. All  $H^\infty$  optimal aposteriori adaptive filters that achieve  $\gamma_{opt} = 1$  are given by

$$\begin{aligned} \hat{z}_{i|i} &= h_i \hat{x}_{i|i} + (I + \mu |h_i|^2)^{-\frac{1}{2}} \\ &\quad S_i \left( \{ (I + \mu |h_j|^2)^{\frac{1}{2}} (d_j - h_j \hat{x}_{j|j}) \}_{j=0}^i \right), \end{aligned}$$

where  $\hat{x}_{i|i}$  is the estimated state of the normalized LMS algorithm with parameter  $\mu$ , and  $S(a_i, \dots, a_0)$  is any (possibly nonlinear) contractive causal mapping as described before.

The choice  $S = 0$  yields the normalized LMS, whereas  $S = I$  (the identity map) yields  $\hat{z}_{i|i} = d_i$ , the estimator in (8).

If the input data  $\{h_i\}$  is exciting, and the bound on  $\mu$  is satisfied, then all  $H^\infty$  optimal apriori adaptive filters are given by

$$\begin{aligned} \hat{z}_i &= h_i \hat{x}_i + (I - \mu |h_i|^2)^{\frac{1}{2}} \\ &\quad S_i \left( \{ (I - \mu |h_j|^2)^{\frac{1}{2}} (d_j - h_j \hat{x}_j) \}_{j=0}^{i-1} \right), \end{aligned}$$

where  $\hat{x}_i$  is the state estimate of the LMS algorithm with learning rate  $\mu$ , and  $S$  is any (possibly nonlinear) contractive causal mapping.

As before,  $S = 0$  yields LMS. However, the choice  $S = I$  yields the highly nontrivial estimator  $\hat{z}_i = h_i \hat{x}_i + (I - \mu |h_i|^2)^{\frac{1}{2}} (I - \mu |h_{i-1}|^2)^{\frac{1}{2}} (d_{i-1} - h_{i-1} \hat{x}_i)$ .

### VII. RISK SENSITIVE OPTIMALITY

We now focus on a certain property of the central  $H^\infty$  filters, namely the fact that they are risk-sensitive optimal filters (see, *e.g.*, [11]). This will give a stochastic interpretation for the LMS algorithm in the special case of disturbances that are independent Gaussian random variables.

#### Theorem 5 (Normalized LMS)

Consider the state-space model (6) where the  $w$  and  $\{v_j\}$  are now assumed independent Gaussian random variables with means  $\hat{w}_{i-1}$  and 0, and variances

$\mu I$  and  $I$ , respectively. The solution to the following minimization problem

$$\min_{\{z_{j,i}\}} \mu_{f,i}(\theta) = \min_{\{z_{j,i}\}} \left( 2 \log \left[ E \exp \left( \frac{1}{2} C_{f,i} \right) \right] \right)$$

where  $C_{f,i} = \sum_{j=0}^i e_{f,i}^* e_{f,i}$ , and the expectation is taken over  $w$  and  $\{v_j\}$  subject to observing  $\{d_0, d_1, \dots, d_i\}$ , is given by the normalized LMS algorithm with parameter  $\mu$ .

**Theorem 6 (LMS)** Consider the same model (6) where the  $w$  and  $\{v_j\}$  are assumed independent Gaussian random variables with means  $w_{i-1}$  and 0, and variances  $\mu I$  and  $I$ , respectively. Suppose, moreover, that the  $\{h_i\}$  are exciting and that the bound on  $\mu$  is satisfied. Then the solution to the following minimization problem

$$\min_{\{z_j\}} \mu_{p,i}(\theta) = \min_{\{z_j\}} \left( 2 \log \left[ E \exp \left( \frac{1}{2} C_{p,i} \right) \right] \right)$$

where  $C_{p,i} = \sum_{j=0}^i e_{p,i}^* e_{p,i}$ , and the expectation is taken over  $w$  and  $\{v_j\}$  subject to observing  $\{d_0, d_1, \dots, d_{i-1}\}$ , is given by the LMS algorithm with learning rate  $\mu$ .

Some intuition regarding this result can be gained by elaborating on the above cost functions. These are convex and increasing functions in  $C_{f,i}$  and  $C_{p,i}$ . Such a criterion is termed *risk-averse* (or *pessimistic*) since large weights are on large values of  $C_{f,i}$  (or  $C_{p,i}$ ), and hence we are more concerned with the occasional occurrence of large values than with the frequent occurrence of moderate ones. Thus LMS and normalized LMS are risk-averse filters that avoid the occasional occurrence of large estimation error energies, at the expense of admitting the frequent occurrence of moderate values of estimation error energy.

## VIII. CONCLUDING REMARKS

We have demonstrated that the LMS algorithm is  $H^\infty$  optimal. This result solves a long standing issue of finding a rigorous basis for the LMS algorithm, and also confirms its robustness.

## Acknowledgement

The authors would like to thank Prof. L. Ljung for contributing to the discussion in Section V.4.

## REFERENCES

- [1] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," *IRE WESCON Conv. Rec.*, Pt. 4, pp. 96-104, 1960.
- [2] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, Englewood Cliffs, NJ, 1991.
- [3] P.P. Khargonekar and K. M. Nagpal, "Filtering and smoothing in an  $H^\infty$ -setting," *IEEE Trans. on Automatic Control*, AC-36, pp. 151-166, 1991.
- [4] T. Basar, "Optimum performance levels for minimax filters, predictors and smoothers," *Systems and Control Letters*, vol. 16, pp. 309-317, 1991.
- [5] D. J. Limebeer and U. Shaked, "New results in  $H^\infty$ -filtering," In *Proc. Int. Symp. on MTNS*, pp. 317-322, June 1991.
- [6] U. Shaked and Y. Theodor, " $H^\infty$ -optimal estimation: A tutorial," In *Proc. IEEE Conference on Decision and Control*, pp. 2278-2286, Tucson, AZ, Dec. 1992.
- [7] M. J. Grimble, "Polynomial matrix solution of the  $H^\infty$  filtering problem and the relationship to Riccati equation state-space results," *IEEE Trans. on Signal Processing*, SP-41, no. 1, pp. 67-81, January 1993.
- [8] B. Hassibi, A. H. Sayed, and T. Kailath, "Recursive linear estimation in Krein spaces - Part I: Theory," in *Proc. of this Conference*.
- [9] B. Hassibi, A. H. Sayed, and T. Kailath, "Recursive linear estimation in Krein spaces - Part II: Applications," in *Proc. of this Conference*.
- [10] A. H. Sayed and T. Kailath, "A state-space approach to adaptive filtering," in *Proc. ICASSP*, vol. 3, pp. 559-562, Minneapolis, MN, April 1993.
- [11] P. Whittle, *Risk Sensitive Optimal Control*, John Wiley and Sons, New York, 1990.